



Open access – Research article

THIS ARTICLE IS PART OF A SPECIAL ISSUE ENTITLED
'MOLECULAR TECHNOLOGIES TO IMPROVE
VEGETATIVELY PROPAGATED BANANA AND CASSAVA'A platform for efficient genotyping in *Musa* using
microsatellite markers

Pavla Christelová¹, Miroslav Valárik¹, Eva Hřibová¹, Ines Van den houwe², Stéphanie Channelière³,
Nicolas Roux³ and Jaroslav Doležel^{1*}

¹ Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, Olomouc CZ-77200, Czech Republic

² Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Kasteelpark Arenberg 13, B-3001 Leuven, Belgium

³ Bioversity International, Parc Scientifique Agropolis II, Montpellier Cedex 5, 34397, France

Received: 17 March 2011; **Returned for revision:** 14 May 2011; **Accepted:** 16 August 2011; **Published:** 22 August 2011

Citation details: Christelová P, Valárik M, Hřibová E, Van den houwe I, Channelière S, Roux N, Doležel J. 2011. A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB PLANTS* 2011 plr024 doi:10.1093/aobpla/plr024

Abstract

**Background
and aims**

Bananas and plantains (*Musa* spp.) are one of the major fruit crops worldwide with acknowledged importance as a staple food for millions of people. The rich genetic diversity of this crop is, however, endangered by diseases, adverse environmental conditions and changed farming practices, and the need for its characterization and preservation is urgent. With the aim of providing a simple and robust approach for molecular characterization of *Musa* species, we developed an optimized genotyping platform using 19 published simple sequence repeat markers.

Methodology

The genotyping system is based on 19 microsatellite loci, which are scored using fluorescently labelled primers and high-throughput capillary electrophoresis separation with high resolution. This genotyping platform was tested and optimized on a set of 70 diploid and 38 triploid banana accessions.

Principal results

The marker set used in this study provided enough polymorphism to discriminate between individual species, subspecies and subgroups of all accessions of *Musa*. Likewise, the capability of identifying duplicate samples was confirmed. Based on the results of a blind test, the genotyping system was confirmed to be suitable for characterization of unknown accessions.

Conclusions

Here we report on the first complex and standardized platform for molecular characterization of *Musa* germplasm that is ready to use for the wider *Musa* research and breeding community. We believe that this genotyping system offers a versatile tool that can accommodate all possible requirements for characterizing *Musa* diversity, and is economical for samples ranging from one to many accessions.

* Corresponding author's e-mail address: dolezel@ueb.cas.cz

Introduction

The important role of bananas and plantains (*Musa* spp.) as one of the top world trade commodities and as food security for millions of people, especially in humid tropics, is unquestionable. However, this crop faces serious endangerment by numerous pests and diseases. Breeding efforts are hampered by a high degree of banana sterility and a lack of characterized germplasm as potential parents for breeding. Currently grown banana cultivars are mainly triploid clones, which originated as intraspecific hybrids of *Musa acuminata* and interspecific hybrids between *M. acuminata* and *Musa balbisiana*, with a possible involvement of a few other species within the genus. To set up an efficient strategy for breeding improved banana varieties and support the choice of crossing parents, a solid understanding of the genetic diversity of available resources is needed. Likewise, conservation of existing gene resources is essential, especially when we observe the continuous loss of banana diversity due to indelicate environmental treatment of the rain forests, as well as changed farming practices of smallholders. The main objectives and means for *Musa* diversity conservation were formulated in the Global Conservation Strategy for *Musa* (INIBAP 2006) under the scope of GMGC (Global Musa Genomics Consortium). Nevertheless, irrespective of the selected strategy, efficient collection and preservation of banana diversity highly depend on unambiguous sample identification. To avoid problems of duplicates within national, regional and global germplasm collections, an accurate and standardized characterization of newly introduced accessions as well as those already deposited in gene banks would be of great benefit. This rationalization effort will allow *Musa* accessions to be efficiently conserved.

Traditional classification of *Musa* species is based on morphological characters and chromosome counts (basic chromosome number; x) (Cheesman 1947; Simmonds and Shepherd 1955). Although a morpho-taxonomic system allows for differentiation of specific banana clones (Stover and Simmonds 1987), insufficiencies of this approach start to emerge as the genetic basis of the plants under study gets narrow. Additionally, a small change at the DNA level can cause a large phenotypic manifestation, while sometimes no or minor morphological changes can be observed after extensive genetic changes. Obviously, a classification system that relies exclusively on the phenotypic manifestations of the genome suffers from limited accuracy (Crouch et al. 2000; De Langhe et al. 2005), but can be made robust if supported by molecular-based characterization.

The enormous increase in the availability of various molecular techniques over the past decades has facilitated the classification of new banana cultivars, as well as reassessment of the traditional taxonomy. Among the broad portfolio of molecular tools, some of the markers have gained special attention in terms of their use in diversity studies and molecular characterization of banana genotypes. Most recently, diversity arrays technology was used for the assessment of genetic diversity within *Musa* spp. (Risterucci et al. 2009). While having the advantage of a high-throughput approach suitable for large numbers of genotypes, its use for a limited number of samples in a short turn-around time would rank it within the more demanding methods in terms of funding support. The same applies to the genotyping by sequencing approach, which has gained special attention recently (Elshire et al. 2011). Other molecular markers applied in *Musa* diversity studies were RAPDs (random amplified polymorphic DNA; Pillay et al. 2000, 2001; Ruangsuttapha et al. 2007; Venkatachalam et al. 2008) and AFLPs (amplified fragment length polymorphisms; Loh et al. 2000; Wong et al. 2001a; Ude et al. 2002; Wang et al. 2007). Both these markers have a relatively high level of polymorphism, but they are dominant and, in the case of RAPDs, their reproducibility is a serious limitation (Jones et al. 1997). The more advantageous co-dominant markers were also used for *Musa*, such as RFLPs (restriction fragment length polymorphisms; Gaweł et al. 1992; Nwakanma et al. 2003; Ning et al. 2007) and SSRs (simple sequence repeats; e.g. Kaemmer et al. 1997; Grapin et al. 1998; Lagoda et al. 1998; Buhariwalla et al. 2005). While RFLPs perform well in terms of reproducibility, they have a relatively low level of polymorphism and are difficult to use. On the contrary, SSR markers outperform the RFLPs and RAPDs in all the above-mentioned aspects.

Microsatellites (SSRs) are stretches of simple 1- to 6-base-pair-long repeat motifs arranged tandemly within the genomes of prokaryotic and eukaryotic organisms. Their flanking regions, which are usually highly conserved, are suitable for designing locus-specific primers. Simple sequence repeats have been successfully applied in the molecular genotyping of many important crops such as rice (Pessoa-Filho et al. 2007), cereals (Hayden et al. 2007), grapevine (This et al. 2004) or cacao (Zhang et al. 2006). Moreover, the use of SSR markers opens up the possibility of automation and multiplexing, which significantly increases the throughput of the technique.

With the aim of developing a standardized protocol to classify *Musa* germplasm, we have tested and optimized the use of 22 published SSR markers on a set of banana

genotypes. The goal of the present study was to investigate the potential of this marker set to distinguish individual accessions and to develop a standardized procedure for *Musa* genotyping that could serve as a basis for molecular characterization of new samples introduced into the global *Musa* gene bank (International Transit Centre (ITC), Leuven, Belgium) as well as to the wider *Musa* research and breeding community.

Materials and methods

Plant material and the reference DNA collection

The reference DNA collection, comprising a total of 65 accessions [Additional Information 1], was established to represent genetic diversity within the genus *Musa*. *In vitro* plantlets of these accessions are available for distribution from the Bioversity ITC. The genomic DNA of 61 of the 65 accessions is stored in the Genome Resources Centre (http://www.musagenomic-s.org/cetest_firstpage1/genomic_dna.html) and is available for distribution. Out of the 65 accessions, 54 were successfully included in the analysis [Additional Information 1]. To extend the diploid representation of the genotype set, 39 additional diploid accessions were included [Additional Information 2], with three of them being duplicate samples to the Reference DNA collection. These duplicates were included intentionally to test the capability of the genotyping platform to identify sample duplicates. All 39 additional diploid accessions originated from the ITC collection (Leuven, Belgium) as *in vitro* rooted plants and were maintained in a heated greenhouse after transfer to soil. The DNA of these 39 entries was isolated from young leaf tissue using the Invisorb® Spin Plant Mini kit (Invitex, Berlin, Germany), following the manufacturer's instructions.

Polymerase chain reaction amplification and fragment analysis

The 22 SSR loci (Table 1) were amplified using specific primers (Crouch et al. 1998; Lagoda et al. 1998; Hippolyte et al. 2010) that were adjusted by 5'-M13 tails to enable the use of universal fluorescently labelled primer according to Schuelke (2000). Four different fluorophores were used for the primer labelling [6-carboxyfluorescein (6-FAM), VIC, NED and PET; Applied Biosystems, Foster City, CA, USA], allowing for subsequent multiplexing of the reactions (Table 1). The reaction was performed in a final volume of 20 µL containing 10 ng of template genomic DNA, reaction buffer (consisting of 10 mM Tris-HCl (pH 8), 50 mM KCl, 0.1% Triton-X100 and 1.5 mM MgCl₂), 200 µM dNTPs (each), 1 U of *Taq* polymerase, 8 pmol of the M13-tailed locus-specific forward primer, 6 pmol of the fluorescently labelled

universal M13 forward primer and 10 pmol of the locus-specific reverse primer. The cycling conditions were set as follows: initial denaturation step at 94 °C for 5 min, followed by 35 cycles of denaturation (94 °C/45 s), annealing at the temperature corresponding to the locus-specific primer (1 min) and extension (72 °C/1 min). Final extension was allowed for 5 min at 72 °C. The polymerase chain reaction (PCR) products were purified by ethanol/sodium acetate precipitation. Three independent PCR reactions were performed in order to improve the accuracy of allele binning.

For automatic capillary electrophoresis, optimized amounts of amplification products were combined with highly deionized formamide and internal standard (GeneScan™-500 LIZ size standard; Applied Biosystems). After 5 min denaturation at 95 °C, samples were loaded onto the automatic 96-capillary ABI 3730xl DNA Analyzer, and electrophoretic separation and signal detection were carried out with default module settings. In order to reduce the cost and increase the capacity of the genotyping platform, samples were multiplexed for the second and third round of electrophoretic separation. Up to 4-fold multiplexing was applied by combining four PCR products, labelled with different fluorescent dyes (6-FAM, VIC, NED and PET; Table 1) into a single sample for loading. The level of multiplexing could be further increased by combining products of different expected lengths, labelled with the identical fluorescent dyes.

Fragment sizing and data analysis

The resulting data were analysed using GeneMarker® v1.75 (Softgenetics, LLC, State College, PA, USA). Automated scoring of the data was followed by a careful manual check, and low-quality DNA samples were discarded from the analysis. The marker panels were built based on allele calls of the Reference DNA collection sample set and later extended by additional diploid accession allele calls, in order to increase the reference SSR-profiles database. Bins for each allele were set with respect to the allele frequencies and signal strength extracted from the three repeated runs of each sample.

The diploid and triploid accessions were analysed separately, because in the case of polyploid species, the polysomic inheritance brings the simultaneous occurrence of several alleles of a single SSR. In such a situation, the exact number of copies of individual alleles cannot be determined; therefore, the genotypic data are converted into binary data (coded by 1—presence/0—absence) and analysed as a dominant marker's record (Weising et al. 2005). Both genotypic and binary data were used to generate genetic similarity matrices based on Nei's genetic distance coefficient

Table 1 Detailed list of the SSR markers used in the study.

Marker	Fluorophore	Motif	Reference	Accession GenBank	Annealing temperature (this study; °C)	Minimum allele (this study; bp)	Maximum allele (this study; bp)
mMaCIR01	6-FAM	(GA)20	Lagoda et al. (1998)	X87262	55	241	440
mMaCIR03	6-FAM	(GA)10	Lagoda et al. (1998)	X87263	55	111	147
mMaCIR07	NED	(GA)13	Lagoda et al. (1998)	X87258	53	136	195
mMaCIR08	VIC	(TC)6N24(TC)7	Lagoda et al. (1998)	X87264	55	229	283
mMaCIR13	PET	(GA)16N76(GA)8	Lagoda et al. (1998)	X90745	53	268	427
mMaCIR24	PET	(TC)7	Lagoda et al. (1998)	Z85972	48	240	291
mMaCIR27 ^a	PET	(GA)9	Lagoda et al. (1998)	Z85962	58	232	277
mMaCIR39	VIC	(CA)5GATA(GA)5	Lagoda et al. (1998)	Z85970	52	329	390
mMaCIR40	6-FAM	(GA)13	Lagoda et al. (1998)	Z85977	54	169	247
mMaCIR45	6-FAM	(TA)4CA(CTCGA)4	Lagoda et al. (1998)	Z85968	57	274	318
mMaCIR150	VIC	(CA)10	Hippolyte et al. (2010)	AM950440	54	253	376
mMaCIR152	6-FAM	(CTT)18,(CT)17,(CA)6	Hippolyte et al. (2010)	AM950442	54	147	195
mMaCIR164	VIC	(AC)14	Hippolyte et al. (2010)	AM950454	55	256	458
mMaCIR195 ^a	VIC	(GA)11,(GA)6	Hippolyte et al. (2010)	AM950461	54	262	306
mMaCIR196	NED	(TA)4, (TC)17, (TC)3	Hippolyte et al. (2010)	AM950462	55	163	201
mMaCIR214	NED	(AC)7	Hippolyte et al. (2010)	AM950480	53	115	238
mMaCIR231	NED	(TC)10	Hippolyte et al. (2010)	AM950497	55	236	286
mMaCIR260	PET	(TG)8	Hippolyte et al. (2010)	AM950515	55	204	264
mMaCIR264	6-FAM	(CT)17	Hippolyte et al. (2010)	AM950519	53	234	383
mMaCIR307	NED	(CA)6	Hippolyte et al. (2010)	AM950533	54	143	173
Ma-1-32 ^a	NED	(GA)17AA(GA)8AA(GA)2	Crouch et al. (1998)	n/a	58	208	251
Ma-3-90	PET	(CT)11	Crouch et al. (1998)	n/a	53	147	191

^aExcluded from the analysis due to unreproducible amplification.

([Nei 1973](#)) in the software PowerMarker v3.25 ([Liu and Muse 2005](#)). The unweighted pair-group method with arithmetic mean (UPGMA; [Michener and Sokal 1957](#)) was used to assess the relationship between individual genotypes. The results of UPGMA cluster analysis were visualized in the form of a tree using TreeView v1.6.6 ([Page 1996](#)). Polymorphism information content (PIC) and heterozygosity of individual markers were estimated in PowerMarker v3.25. The overall probability of identity (P_{ID}) of unrelated multilocus genotypes was assessed according to [Paetkau et al. \(1995\)](#), as implemented in the IDENTITY program ([Wagner and Sefc 1999](#)).

Blind test

In order to verify the reliability of the optimized genotyping platform and its potential as a standardized

methodology for molecular characterization of new accessions, a set of anonymous samples was analysed [Additional Information 3]. The genomic DNA was extracted from lyophilized leaf tissue provided by the ITC, and samples were analysed following an identical experimental procedure as for the reference DNA collection. Negative and positive controls (five previously analysed reference genotypes) were included in the blind test to ensure correct allele sizing and control the consistency of the electrophoretic condition. The unknown samples were coded numerically and their true identity was disclosed by our partners only after the data analysis. As revealed subsequently, the blind test sample set contained an additional four samples that were duplicates of the reference DNA collection [see Additional Information 1 and 3].

Genotyping error handling

To eliminate genotyping errors, several precautions were employed in the genotyping process, following the recommendations by Bonin et al. (2004). First, to minimize the allelic dropout effect, the multitube approach (Taberlet et al. 1996) was used with three independent reactions for each marker/genotype combination. The error-prone samples with low-quality DNA were discarded from the analysis. Second, the multilocus genotype was examined and accessions differing at a single locus were carefully inspected and reanalysed (if needed) to confirm the difference. Third, to decrease human factor errors, sample preparation was performed by two different people for the replicated reactions. Data evaluation was ruled by strictly pre-set parameters to avoid errors such as misinterpretation of stutter peaks.

Results

Twenty-two SSR markers were selected by CIRAD as a set enabling one to distinguish between individuals in the *Musa* reference DNA collection (Crouch et al. 1998; Lagoda et al. 1998; Hippolyte et al. 2010; Website 1; Table 1). After the initial double-repeated primer test screening using our protocol, 19 markers were selected out of the initial 22 markers set, for their clear reproducible amplification pattern. The three markers that were excluded from the analysis produced extensive stuttering of peaks, disabling the reproducible interpretation of the SSR profiles. All further analyses were performed with the selected 19 SSRs. Altogether, the SSR profiles were collected for 70 diploid and 38 triploid banana accessions. All necessary information on the genotyping methodology as well as the complete allele score files for the analysed genotypes are also available online through <http://olomouc.ueb.cas.cz/musa-genotyping-centre>.

Analysis of diploid accessions

Diploid accessions were underrepresented in the reference DNA collection; therefore, we decided to include additional diploids in the analysis to increase the number of reference SSR profiles [Additional Information 2]. In the resulting set of 70 diploid accessions (including the blind test entries), a total of 292 alleles were scored from the 19 loci, with an average of 15.4 alleles per locus. The observed heterozygosity (the fraction of all individuals who are heterozygous for the observed locus) ranged between 0.179 and 0.714 (mean 0.450). The PIC of the markers used was relatively high (mean 0.827), ranging between 0.625 and 0.936 (see Table 2 for details). The P_{ID} (combined over all loci), which represents the probability of observing identical genotypes

purely by chance, was 9.44×10^{-29} , denoting the extremely high resolution power of this marker set.

The UPGMA cluster analysis based on the Nei (1973) genetic distance revealed a relatively clear grouping of genotype groups and subgroups (Fig. 1). The B-genome representatives *M. balbisiana* including the diploid hybrid cultivars (AB and BB×T) formed a separate cluster (cluster I). The A-genome representatives *M. acuminata* species were grouped in several clusters depending on their subspecies classification. *Musa acuminata* ssp. *banksii* entries grouped within cluster II, *M. acuminata* ssp. *microcarpa* grouped together with *Musa schizocarpa* and AS hybrids within cluster III. The sole representative of *errans* subspecies, cultivar Agutay, was present at the separate clade related to the above-described *M. acuminata* clusters. Subcluster VI contained the *M. acuminata* ssp. *zebrina* representatives. Subspecies *burmannica*, *burmannicoides* and *siamea* were grouped within cluster VII, sharing their position with several entries from the section *Rhodochlamys*. *Musa acuminata* ssp. *malaccensis* subspecies formed a separate cluster labelled VIII (Fig. 1). Most of the AA cultivars were grouped within cluster IV. The Australimusa section representatives included in the study formed cluster V, together with *Musa beccarii* (classified under the *Callimusa* section). *Musa coccinea*, another representative of the *Callimusa* section, was separated from all the other groups, resembling the behaviour of an outgroup species. As mentioned before, *Rhodochlamys* species were partly present in cluster VII (specifically the *Musa ornata* and *Musa mannii* entries). *Musa velutina* accessions, another representative of the *Rhodochlamys* section, formed a separate cluster labelled IX together with a single *M. ornata* accession (ITC 1330).

Blind test with diploid accessions

When the anonymous samples were included in the dataset, the clustering was slightly changed (Fig. 2). The position of accession Agutay (*M. acuminata* ssp. *errans*) moved into cluster II containing mostly the *M. acuminata* ssp. *banksii* entries. Another alteration could be seen in the position of *M. acuminata* ssp. *zebrina* species, which no longer formed a separate subclade (previously labelled with VI), but instead clustered within cluster IV containing the AA cultivars. Finally, cluster VII, although not changed in the content, now showed a different subclustering pattern, with the *M. acuminata* ssp. *burmannica*, *burmannicoides* and *siamea* species grouped together within one subcluster (VIIa), separated from the *Rhodochlamys* entries (subcluster VIIb).

Out of the nine anonymous accessions, eight were assessed correctly as the closest related species to the corresponding reference accession (Fig. 2). The only

Table 2 Allele number, frequency of the major allele, unique genotypes observed, heterozygosity and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 70 diploid *Musa* accessions.

Marker	Major allele frequency	Number of unique genotypes observed	Allele number	Observed heterozygosity	PIC ^a
mMaCIR01	0.125	39	26	0.531	0.936
mMaCIR03	0.357	13	7	0.400	0.694
mMaCIR07	0.181	33	21	0.551	0.883
mMaCIR08	0.231	22	12	0.646	0.830
mMaCIR13	0.229	28	19	0.543	0.870
mMaCIR24	0.328	19	15	0.344	0.767
mMaCIR39	0.200	39	20	0.714	0.893
mMaCIR40	0.233	29	23	0.534	0.887
mMaCIR45	0.207	16	8	0.357	0.801
mMaCIR150	0.328	20	15	0.522	0.797
mMaCIR152	0.232	19	11	0.250	0.849
mMaCIR164	0.161	28	22	0.322	0.916
mMaCIR196	0.250	23	13	0.453	0.855
mMaCIR214	0.383	12	7	0.313	0.670
mMaCIR231	0.214	27	14	0.540	0.880
mMaCIR260	0.329	20	14	0.357	0.765
mMaCIR264	0.239	35	24	0.522	0.900
mMaCIR307	0.500	10	6	0.179	0.625
Ma-3-90	0.167	31	15	0.474	0.893
Mean	0.258	24.4	15.4	0.450	0.827

^aPolymorphism information content.

exception was blind sample no. 4 (*M. acuminata* ssp. *malaccensis* ITC 0250), which did not group together with its reference genotype (the same ITC 0250 accession), but instead clustered together within the *M. acuminata* ssp. *banksii* subgroup (clade II). The multi-locus genotypes of the blind sample no. 4 (ITC 0250) and the closest related genotype Higa (ITC 0428) differed at a single locus only, suggesting that the blind sample no. 4 belonged very likely to the *banksii* subspecies.

In order to further investigate this incongruence in the blind test results, we conducted internal transcribed spacer (ITS) locus sequence analysis according to Hřibová et al. (2011) in the problematic *malaccensis* accessions. This analysis confirmed that the blind sample no. 4 was not identical to the genotype *M. acuminata* ssp. *malaccensis* ITC 0250, which was originally received from the ITC and stored in the local greenhouse [see Additional Information 4]. The results are, however, not conclusive about the identity of blind sample no. 4, as only a single representative of the

banksii subspecies was used for the ITS analysis in our previous study. Thus, it cannot be explicitly stated whether blind sample no. 4 is a different genotype of *M. acuminata* ssp. *malaccensis* or ssp. *banksii*, or rather a hybrid between *malaccensis* and *banksii* subspecies. Only a more detailed sequence analysis would probably provide a definite answer.

Analysis of triploid accessions

Altogether, 38 triploid accessions were analysed (including the blind test entries). The 19 microsatellite loci scored a total of 267 alleles, ranging between 8 and 24 per locus, with a mean value of 14 alleles per locus. The average PIC of the SSR markers applied on the triploid accessions was 0.850 (Table 3).

The UPGMA analysis majority rule consensus tree showed two main clusters, cluster A and cluster B (Fig. 3). Cluster A contained solely the AAA hybrid accessions, with a separated clade bearing the Lujugira/Mutika subgroup representatives, as well as a

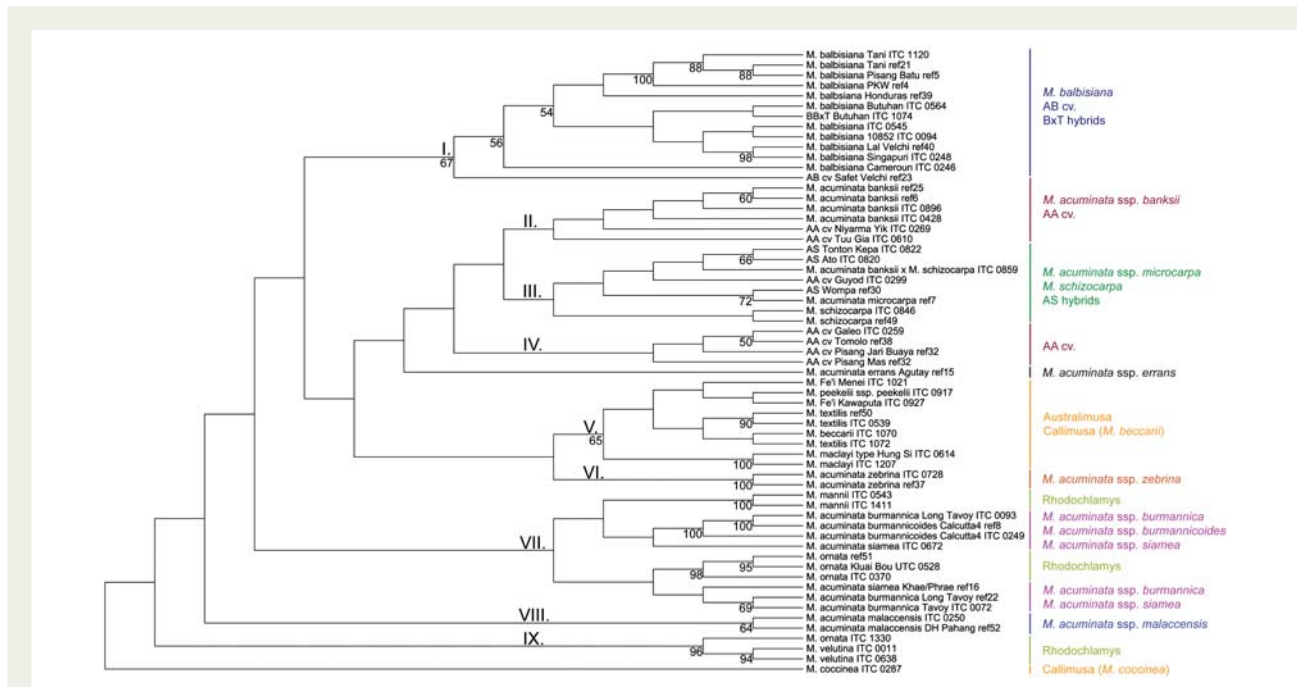


Fig. 1 Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset. Bootstrap support values higher than 50% are marked below the corresponding branches. The classification of the genotypes into individual sections, species and subspecies of the genus *Musa* is indicated by the coloured side bars and legends. A complete list of accessions with their taxonomic details can be found in [Additional Information 1 and 2].

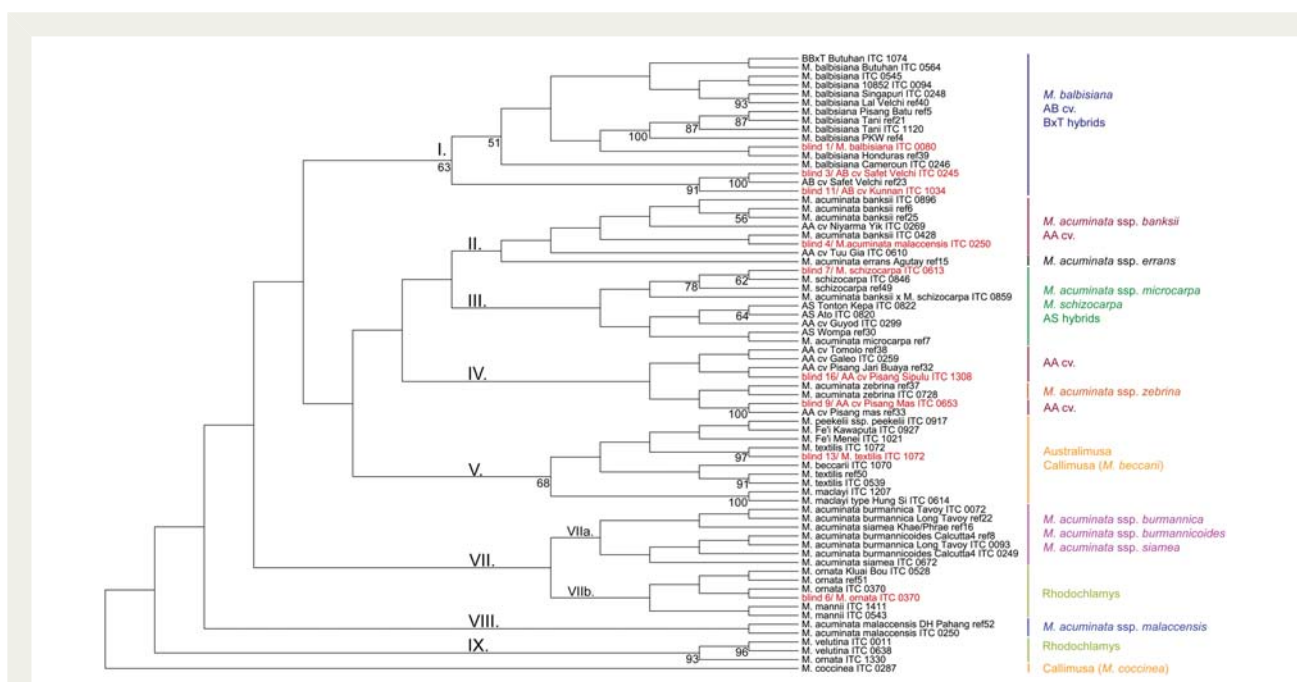


Fig. 2 Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset including the blind test samples. Bootstrap support values higher than 50% are marked below the corresponding branches. The anonymous samples included in the blind test are highlighted in red. The classification of the genotypes into individual sections, species and subspecies of the genus *Musa* is indicated by the coloured side bars and legends. A complete list of accessions with their taxonomic details can be found in [Additional Information 1, 2 and 3].

Table 3 Major allele frequency, allele number and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 38 triploid *Musa* accessions.

Marker	Major allele frequency	Allele number	PIC
mMaCIR01	0.105	24	0.942
mMaCIR03	0.237	12	0.839
mMaCIR07	0.132	17	0.912
mMaCIR08	0.237	14	0.867
mMaCIR13	0.342	12	0.804
mMaCIR24	0.289	12	0.817
mMaCIR39	0.316	18	0.859
mMaCIR40	0.289	9	0.817
mMaCIR45	0.289	12	0.814
mMaCIR150	0.263	8	0.808
mMaCIR152	0.263	12	0.850
mMaCIR164	0.131	18	0.913
mMaCIR196	0.237	15	0.881
mMaCIR214	0.263	8	0.788
mMaCIR231	0.132	16	0.905
mMaCIR260	0.474	13	0.733
mMaCIR264	0.158	18	0.913
mMaCIR307	0.342	8	0.760
Ma-3-90	0.105	21	0.934
Mean	0.242	14.1	0.850

distinct clade leading to the edible species from the Cavendish and Gros Michel subgroups. Among all the AAA entries included in the analysis, only the accession Pisang Berangan clustered outside the A cluster, sharing a clade (IVa) with the African plantain representatives within the main cluster B. The second main cluster B was split into four subclusters/subclades. While subcluster II was formed exclusively by the AAB hybrid entries, subcluster I also contained an ABB genotype Namwa Khom (Pisang Awak subgroup), as a closest relative of the AAB Figue Pomme Géante accession from the Silk subgroup. Two of the ABB hybrid representatives, Kluai Tiparot and Pelipita, formed the third subclade within the B cluster (III). Most of the ABB hybrids were grouped under IVb, together with an AAB accession Popoulou. The African plantains formed a separate clade IVa with a single AAA representative P. Berangan, as mentioned above.

Blind test with triploid accessions

Six encoded triploid samples were included in the blind test and all of them were assessed correctly as the closest related species to the corresponding reference genotype from identical subgroups, with significant statistical support (Fig. 4). The position of some clades was slightly altered after the inclusion of anonymous samples in the analysis (Fig. 4). Specifically, the UPGMA cluster analysis has now shown an altered position of the clade previously labelled III (ABB accessions Pelipita and Kluai Tiparot) and the subclade of the cluster previously labelled II, bearing the AAB genotypes P. Palembang, P. Rajah and P. Raja Bulu. However, the bootstrap statistical support for nodes leading to these clades was not significantly strong in either dataset, and the position of all the other clades in the consensus tree remained unchanged.

Identification of duplicate accessions

One hundred per cent similarity in multilocus genotypes was seen in nine pairs of duplicate accessions [Additional Information 5]. Some of the duplicates were introduced into the accession set intentionally from the local greenhouse (originally coming from the ITC collection) to assess the capability of our genotyping system at spotting the duplicate accessions. Others were introduced through the blind test samples (see Materials and methods). All the duplicates were identified [Additional Information 5], with two exceptions. The *Musa textilis* reference collection DNA sample (ref. 50), which was reported to correspond to the ITC accession ITC 1072, was shown to be identical to another *M. textilis* accession (ITC 0539). This suggests that the reference sample (ref. 50) was mislabelled or its origin was not reported correctly.

Another anticipated duplicate, introduced into the triploid entries through the blind test, was accession blind 12 (Pisang Bakar ITC 1064). Its corresponding reference DNA sample was ref. 19. However, their identity based on the multilocus molecular profile was not approved. Although the two samples differed at 7 out of the 19 scored SSR loci, their closest relationship was revealed after the UPGMA cluster analysis (Fig. 4), suggesting that their mutual subgroup classification (subgr. Ambon) may be correct, but the identity of one of the samples was confused.

Moreover, more than one duplicate accession was reported for both accession ref. 8 (*M. acuminata* ssp. *burmannicoides* 'Calcutta4') and ref. 21 (*M. balbisiana* 'Tani'). The second duplicate for each of the two reference samples was classified under the same species/subspecies [Additional Information 5]. This indicates that

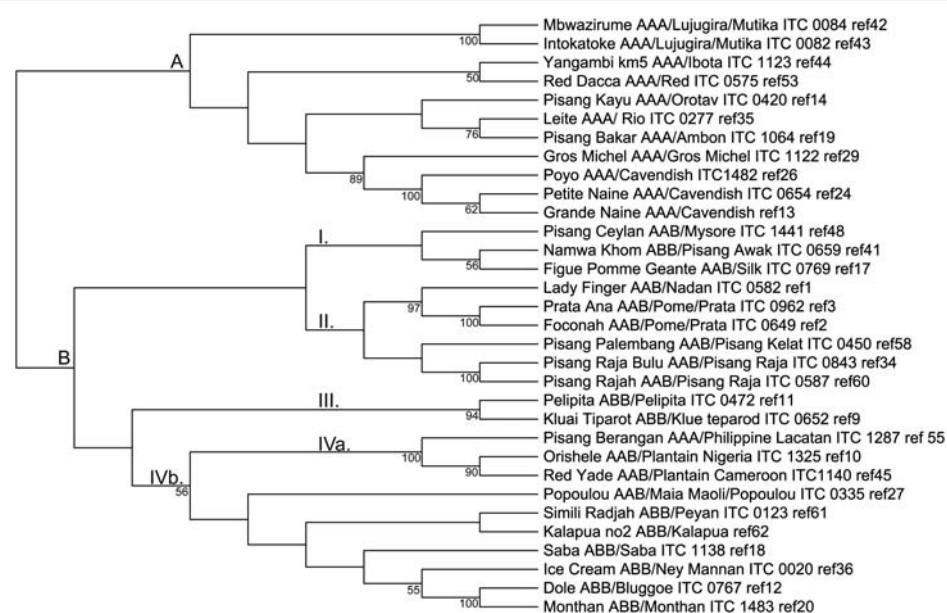


Fig. 3 Dendrogram showing the results of the UPGMA analysis of triploid accessions dataset. Bootstrap support values higher than 50% are marked below the corresponding branches. A complete list of accessions with their taxonomic details can be found in [Additional Information 1 and 2].

either the marker set used did not have enough resolution power to distinguish these accessions or, more likely, based on the low P_{ID} value mentioned above, these accessions were mislabelled.

Discussion

While the use of microsatellite markers to analyse genetic diversity among *Musa* species is well documented (e.g. Kaemmer et al. 1997; Grapin et al. 1998; Buhariwalla et al. 2005; Ning et al. 2007; Venkatachalam et al. 2008; Wang et al. 2010), its application in the form of a standardized platform to serve for genotyping purposes for the wider *Musa* community is still missing. In this study, we attempted to develop an optimized SSR-based system for molecular characterization of *Musa* accessions that could be used as the basis for the foundation of the *Musa* Genotyping Centre (MGC).

Mislabelling of accessions and sample duplications are common problems in germplasm collections (e.g. Virk et al. 1995; Zhang et al. 2006). The resolution of the marker set tested in this study was high enough ($P_{ID} = 9.44 \times 10^{-29}$) to distinguish between different accessions and proved to be powerful enough to identify mislabelled accessions, as documented in the case of the *M. acuminata* ssp. *malaccensis* accession. Similarly, its potential for identifying duplicates was clearly

proved on the present dataset. Nevertheless, we wanted to ensure reproducibility of results and minimization of genotyping errors prior to its implementation into practice. When compared with the original data reported by Lagoda et al. (1998) for a subset of markers, the allele size ranges were overlapping, but not identical. Similar problems have been described previously, and most often they were attributed to the method used and the conditions of electrophoretic separation (e.g. Testolin et al. 2000; Creste et al. 2003). Also, the automatic capillary electrophoresis system used in this experiment allows for much higher resolution and run-to-run precision than the previously used gel-based systems. Therefore, the wider range of allele sizes and higher numbers of identified alleles are adding to the resolution power of the marker set rather than restricting the capability of the platform.

Among the common genotyping errors that are responsible for misidentification of a particular genotype, allele dropout and false allele amplification play an important role. Allele dropout is an accidental failure of PCR to amplify one of the alleles present at the heterozygous locus, which produces false homozygous patterns (Pompanon et al. 2005). To deal with this problem, three options have been proposed. The first relies on systematic replication of the genotyping, i.e. a multitube approach, which in most cases would expose

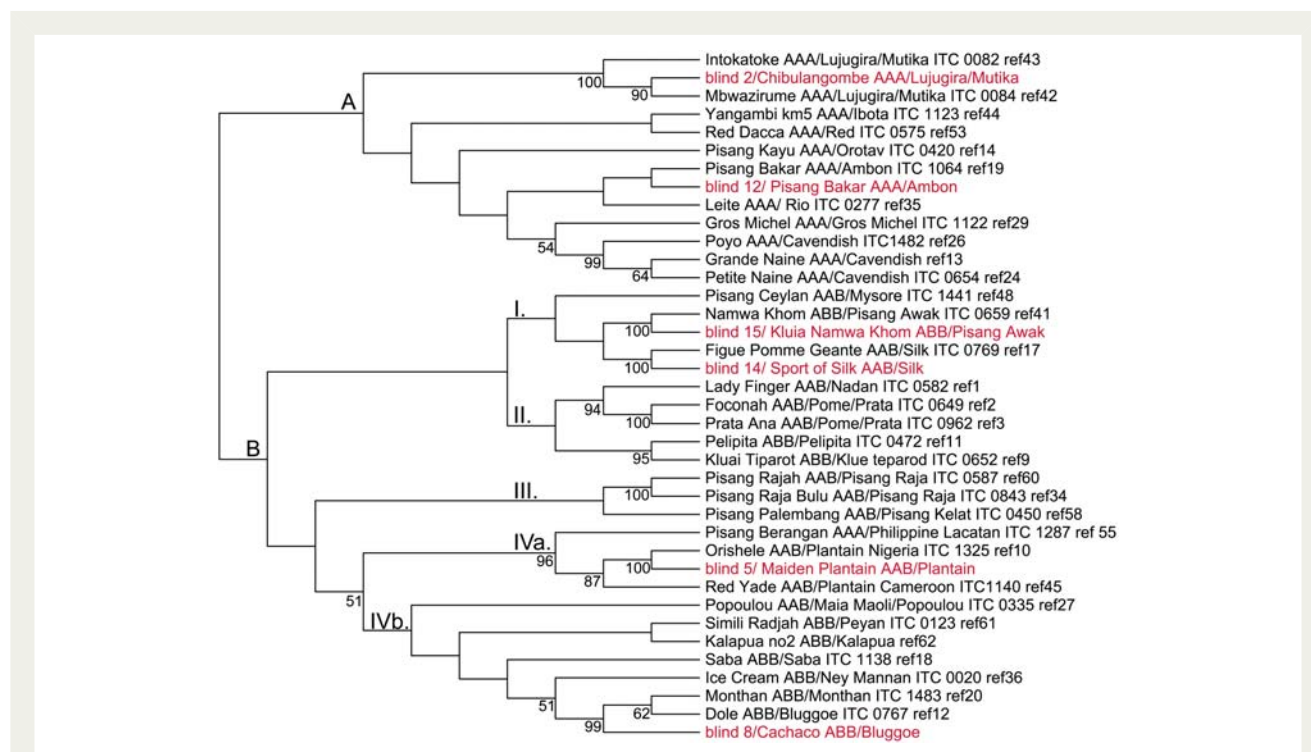


Fig. 4 Dendrogram showing the results of the UPGMA analysis of the triploid accessions dataset including the blind test samples. Bootstrap support values higher than 50% are marked below the corresponding branches. The anonymous samples included in the blind test are highlighted in red. A complete list of accessions with their taxonomic details can be found in [Additional Information 1, 2 and 3].

the underlying allelic dropouts or allele shifts due to poor amplification (Taberlet et al. 1996). Another possibility is to allow for a certain level of mismatch tolerance, provided that enough loci are scored. Then based on the multilocus genotype, the differences generated by genotyping errors can be distinguished from those that are actual differences between two genotypes by the low number of mismatches (McKelvey and Schwartz 2004). The third option combines the two former ones, with replicated genotyping only for samples where three or fewer mismatches at different loci were observed. These multilocus genotypes are re-evaluated after the repeated typing to prove that they are different genotypes in reality, but the cost increase by PCR replications is minimized (Zhang et al. 2006). In this pilot study, we adopted the multitube approach with three replicates to ensure maximum precision. However, with many more samples coming to be analysed in the MGC, and thereby increasing the reference database of molecular profiles, the third (combined) option appears to be adequate and is currently being tested.

The grouping revealed by the UPGMA cluster analysis was consistent with the characterization based on the morphotaxonomic classification of accessions (Figs. 2

and 4). The *Callimusa* section, however, did not form a separate cluster, which reflects its controversial position and agrees with its previously reported close relationship to the *Australimusa* species (Jarret and Gawel 1995; Wong et al. 2001b, 2002). Also, the close relationship between *Rhodochlamys* and *M. acuminata* species (Wong et al. 2002; Bartoš et al. 2005; Li et al. 2010; Liu et al. 2010) was confirmed. The marker set enabled distinction to the level of individual subgroup/subspecies. The degree of polymorphism varied between subgroups and subspecies, and polymorphic sites were still to be found within the subgroups and subspecies. For example, in contrast to the study of Creste et al. (2003) who were not able to find polymorphic loci among the Cavendish subgroup of bananas in their study based on six SSR loci, the marker set used in our study did provide polymorphic loci among the three representatives of the Cavendish subgroup, allowing for their distinction. Obviously, the larger number of loci scored increases the possibility of finding enough polymorphic loci. On the other hand, limitations in the resolution of microsatellite markers become evident when somatic mutants are analysed; as they share the common origin, the genetic variation that is narrowed through

the cycles of vegetative propagation may not be reflected in their SSR molecular profile (Cipriani et al. 1994; Creste et al. 2003; Esselink et al. 2003). As most of the commercial banana cultivars are vegetatively propagated clones, assessment of their genetic variability through the marker set tested in this study may not be successful and is yet to be confirmed. However, it still presents a very useful platform for molecular characterization of unknown samples and assessment of the genetic integrity of the *Musa* germplasm collections.

Although microsatellites have been used as reliable markers for projects with labour division among laboratories (Bredemeijer et al. 2002; Röder et al. 2002), several pieces of work have shown that there was a significant level of incongruence between the results obtained at different workplaces, thus complicating the transferability and comparability of the data (Jones et al. 1997; Weeks et al. 2002; This et al. 2004; Van Treuren et al. 2010). In the light of this, centralization of genotyping activities in *Musa* and its standardization as a service to the research community appear to be preferable options. In addition to facile quality control, the core facility would enable the use of other methods to support the genotyping, such as flow cytometric estimation of ploidy level and/or genome size, keeping in mind that the genotyping data treatment differs for the diploid and polyploid accessions (see Materials and methods). Obviously, sample transfer requirements can be minimized if both types of analysis are performed at a single site. Moreover, with every new sample passing through the analysis, the database of reference SSR profiles is enlarged and the probability of identifying the closest relative or exactly matching accession is enhanced.

Based on our results obtained with the SSR markers presented in this work and those of Hřibová et al. (2011) obtained with ITS, as well as the long-term experience in DNA flow cytometry (Doležel 1991; Lysák et al. 1999; Roux et al. 2003; Bartoš et al. 2005; Doleželová et al. 2005), the MGC has been established at the Institute of Experimental Botany in Olomouc (Czech Republic) under the umbrella of Bioversity International (<http://olomouc.ueb.cas.cz/musa-genotyping-centre>). The Centre serves the whole *Musa* research and breeding community. Moreover, the genotyping platform has already been included in the pipeline for characterization of newly introduced accessions to the international banana germplasm collection (ITC). In this pipeline, fresh leaf tissue samples for molecular characterization are received at the MGC, where they are subjected to ploidy level measurement via flow cytometry; the DNA is extracted and used for collecting the SSR profiles of the 19 markers as described above. In certain cases, where the results of the SSR genotyping are not conclusive enough to reliably classify the

unknown samples, the ITS sequence analysis according to Hřibová et al. (2011) can be applied. Although it is obvious that new high-content, high-throughput, genotyping approaches will gradually replace marker-based systems, we feel confident that the platform described here offers a well-founded and ready-to-use approach, which can be applied immediately and which offers higher flexibility in scaling the analysis with respect to sample size, cost efficiency and turn-around time for results.

Conclusions and forward look

The platform for genotyping of *Musa* germplasm described here provides a robust and reproducible approach to characterize the genetic variability of this important crop, support the management of germplasm collections and direct genotype selection for breeding improved cultivars. The database of molecular profiles keeps growing with every new sample passing through the analytical pipeline, resulting in stepwise improvement in the grouping, and consequently increasing the chance of finding an exact match for unknown samples. As part of the future plans, a batch of tetraploid accessions will be included in the analysis to make it more versatile and satisfying all possible requirements for molecular characterization of the diverse *Musa* gene pool.

Additional information

The following additional information is available in the online version of this article –

File 1: Taxonomic details of the reference DNA collection accessions.

File 2: List of additional diploid accessions from the ITC collection (maintained in a local greenhouse) included in the analysis.

File 3: List of encoded accessions included in the blind test.

File 4: Detailed results of the ITS sequence analysis of blind sample no. 4 and its putative corresponding reference accession—*M. acuminata* ssp. *malaccensis* (ITC 0250).

File 5: List of duplicates identified among the analysed genotypes.

Sources of funding

This work has been supported by Bioversity International (LOA Cfl 2009/48 and LoA Cfl 2010/58), Internal Grant Agency of Palacký University, Olomouc, Czech Republic (grant award no. Prf-2010-001) and by the Ministry of

Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

Contributions by the authors

All authors have contributed to, read and approved the manuscript.

Acknowledgements

We thank our colleague Marie Seifertová for her excellent technical assistance. The work of Pavla Christelová, Miroslav Valárik, Eva Hřibová, Nicolas Roux, Stéphanie Channelière and Jaroslav Doležel was in the context of IAEA Coordinated Research Project 'Molecular Tools for Quality Improvement in Vegetatively Propagated Crops Including Banana and Cassava' (D23027) and the FAO/IAEA Joint Programme.

Conflict of interest statement

None declared.

References

- Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J. 2005. Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. *Cytogenetic and Genome Research* 109: 50–57.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetic studies. *Molecular Ecology* 13: 3261–3273.
- Bredemeijer GMM, Cooke R, Ganai M, Peeters R, Isaac P, Norordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B. 2002. Construction and testing of a microsatellite database containing more than 500 tomato varieties. *Theoretical and Applied Genetics* 105: 1019–1026.
- Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R. 2005. Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molecular Ecology Notes* 5: 327–330.
- Cheesman EE. 1947. Classification of the bananas II. The genus *Musa* L. *Kew Bulletin* 2: 106–117.
- Cipriani G, Frazza G, Peterlunger E, Testolin R. 1994. Grapevine fingerprinting using microsatellite repeats. *Vitis* 33: 211–215.
- Creste S, Neto AT, Silva SD, Figueira A. 2003. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* 132: 259–268.
- Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R. 1998. Segregation of microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Science* 38: 211–217.
- Crouch HK, Crouch JH, Madsen S, Vuylsteke DR, Ortiz R. 2000. Comparative analysis of phenotypic and genotypic diversity among plantain landraces (*Musa* spp., AAB group). *Theoretical and Applied Genetics* 101: 1056–1065.
- De Langhe E, Pillay M, Tenkouano A, Swennen R. 2005. Integrating morphological and molecular taxonomy in *Musa*: the African plantains (*Musa* spp. AAB group). *Plant Systematics and Evolution* 255: 225–236.
- Doležel J. 1991. Flow cytometric analysis of nuclear DNA content in higher plants. *Phytochemical Analysis* 2: 143–154.
- Doleželová M, Doležel J, Van den houe I, Roux N, Swennen R. 2005. Focus on the *Musa* collection: ploidy levels revealed. *Info-Musa* 14: 34–36.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* 6: e19379.
- Esselink GD, Smulders MJ, Vosman B. 2003. Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theoretical and Applied Genetics* 106: 277–286.
- Gawel NJ, Jarret RL, Whitemore AP. 1992. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theoretical and Applied Genetics* 84: 286–290.
- Grapin A, Noyer JL, Carreel F, Dambier D, Baurens FC, Lanaud C, Lagoda P. 1998. Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* 19: 1374–1380.
- Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ. 2007. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Molecular Breeding* 21: 271–281.
- Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, Jenny C, Perrier X, Carreel F, Argout X, Piffanelli P, Khan IA, Miller RNG, Pappas GJ, Mbéguié-A-Mbéguié D, Matsumoto T, De Bernardinis V, Huttner E, Kilian A, Baurens FC, D'Hont A, Cote1 F, Courtois B, Glaszmann JC. 2010. A saturated SSR/DaT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* 10: 65.
- Hřibová E, Čížková J, Christelová P, Taudien S, de Langhe E, Doležel J. 2011. The ITS1–5.8S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. *PLoS ONE* 6: e17863.
- INIBAP. 2006. *Global conservation strategy for Musa (Banana and Plantain)*. A consultative document prepared by INIBAP with the collaboration of numerous partners in the Musa research-and-development community. France: INIBAP.
- Jarret RL, Gawel NJ. 1995. Molecular markers, genetic diversity and systematics. In: Gowen S, ed. *Bananas and plantains*. London: Chapman & Hall, 67–83.
- Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmioli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* 3: 381–390.
- Kaemmer D, Fischer D, Jarret RL, Baurens FC, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda P. 1997. Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* 96: 49–63.
- Lagoda P, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C. 1998. Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* 7: 657–666.

- Li LF, Häkkinen M, Yuan YM, Hao G, Ge XJ. 2010. Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Molecular Phylogenetics and Evolution* 57: 1–10.
- Liu AZ, Kress WJ, Li DZ. 2010. Phylogenetic analyses of the banana family (Musaceae) based on nuclear ribosomal (ITS) and chloroplast (*trnL-F*) evidence. *Taxon* 59: 20–28.
- Liu K, Muse SV. 2005. PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* 21: 2128–2129.
- Loh JP, Kiew R, Set O, Gan LH, Gan YY. 2000. Amplified fragment length polymorphism fingerprinting of 16 banana cultivars (*Musa* cvs.). *Molecular Phylogenetics and Evolution* 17: 360–366.
- Lysák MA, Doleželová M, Horry JP, Swennen R, Doležel J. 1999. Flow cytometric analysis of nuclear DNA content in *Musa*. *Theoretical and Applied Genetics* 98: 1344–1350.
- McKelvey KS, Schwartz MK. 2004. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management* 68: 439–448.
- Michener CD, Sokal RR. 1957. A quantitative approach to a problem of classification. *Evolution* 11: 490–499.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* 70: 3321–3323.
- Ning SP, Xu LB, Lu Y, Huang BZ, Ge XJ. 2007. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. *Scientia Horticulturae* 114: 281–288.
- Nwakanma DC, Pillay M, Okoli BE, Tenkouano A. 2003. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. *Theoretical and Applied Genetics* 108: 154–159.
- Paetkau D, Calvert W, Stirling I, Strobeck C. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4: 347–354.
- Page RDM. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357–358.
- Pessoa-Filho M, Beló A, Alcochete AAN, Rangel PHN, Ferreira ME. 2007. A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. *BMC Plant Biology* 7: 23.
- Pillay M, Nwakanma DC, Tenkouano A. 2000. Identification of RAPD markers linked to A and B genome sequences in *Musa* L. *Genome* 43: 763–767.
- Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A. 2001. Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics* 102: 965–970.
- Pompanon F, Bonin A, Bellemain E, Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6: 847–859.
- Risterucci AM, Hippolyte I, Perrier X, Xia L, Caig V, Evers M, Huttner E, Kilian A, Glaszmann JC. 2009. Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in *Musa*. *Theoretical and Applied Genetics* 119: 1093–1103.
- Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal M. 2002. Construction and analysis of a microsatellite-based database of European wheat varieties. *Theoretical and Applied Genetics* 106: 67–73.
- Roux N, Toloza A, Radecki Z, Zapata-Arias FJ, Doležel J. 2003. Rapid detection of aneuploidy in *Musa* using flow cytometry. *Plant Cell Reports* 21: 483–490.
- Ruangsuttapha S, Eimert K, Schröder MB, Silayoi B, Denduangboripant J, Kanchanapoom K. 2007. Molecular phylogeny of banana cultivars from Thailand based on HAT-RAPD markers. *Genetic Resources and Crop Evolution* 54: 1565–1572.
- Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.
- Simmonds NW, Shepherd K. 1955. Taxonomy and origin of cultivated bananas. *Botanical Journal of the Linnean Society* 55: 302–312.
- Stover RH, Simmonds NW. 1987. *Bananas*, 3rd edn. Burnt Mill, Harlow: Longman Scientific and Technical.
- Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24: 3189–3194.
- Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S. 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 43: 512–520.
- This P, Jung A, Boccaccii P, Borrego J, Botta R, Constantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibañez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlung E, Regner F, Zulini L, Maul E. 2004. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theoretical and Applied Genetics* 109: 1448–1458.
- Ude G, Pillay M, Nwakanma D, Tenkouano A. 2002. Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* 104: 1246–1252.
- Van Treuren R, Kemp H, Ernsting G, Jongejans B, Houtman H, Visser L. 2010. Microsatellite genotyping of apple (*Malus x domestica* Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. *Genetic Resources and Crop Evolution* 57: 853–865.
- Venkatachalam L, Sreedhar RV, Bhagyalakshmi N. 2008. The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars. *Molecular Phylogenetics and Evolution* 47: 974–985.
- Virk PS, Newbury HJ, Jackson TM, Ford-Lloyd BV. 1995. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. *Theoretical and Applied Genetics* 90: 1049–1055.
- Wagner HW, Sefc KM. 1999. *IDENTITY 1.0*. Vienna: Centre for Applied Genetics, University of Agricultural Sciences.
- Wang JY, Zheng LS, Huang BZ, Liu WL, Wu YT. 2010. Development, characterization and variability analysis of microsatellites from a commercial cultivar of *Musa acuminata*. *Genetic Resources and Crop Evolution* 57: 553–563.
- Wang XL, Chiang TY, Roux N, Hao G, Ge XJ. 2007. Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by

- AFLP markers. *Genetic Resources and Crop Evolution* **54**: 1125–1132.
- Website 1.** Global Musa Genomics Consortium. http://www.musagenomics.org/cetest_firstpage1/genomic_dna.html (28 January 2011).
- Weeks DE, Conley YP, Ferrell RE, Mah TS, Gorin MB. 2002.** A tale of two genotypes: consistency between two high-throughput genotyping centers. *Genome Research* **12**: 430–435.
- Weising K, Nybom H, Wollf K, Kahl G. 2005.** *DNA fingerprinting in plants. Principles, methods, and applications*, 2nd edn. London: Taylor & Francis.
- Wong C, Kiew R, Loh JP, Gan LH, Set O, Lee SK, Lum S, Gan YY. 2001a.** Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. *Annals of Botany* **88**: 1017–1025.
- Wong C, Kiew R, Ohn S, Lamb A, Lee SK, Gan LH, Gan YY. 2001b.** Sectional placement of three Bornean species of *Musa* (Musaceae) based on AFLP. *Garden's Bulletin Singapore* **53**: 327–341.
- Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY. 2002.** Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. *Annals of Botany* **90**: 231–238.
- Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA. 2006.** Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Science* **46**: 2084–2092.